

WHAT IS CLAIMED IS:

1. A synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a polypeptide, having a codon composition differing at more than 25% of the codons from a wild type nucleic acid sequence encoding a polypeptide, and having at least 3-fold fewer transcription regulatory sequences relative to the average number of such sequences resulting from random selections of codons at the codons which differ, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites and promoter sequences, and wherein the polypeptide encoded by the synthetic nucleic acid molecule has at least 85% sequence identity to the polypeptide encoded by the wild type nucleic acid sequence.
2. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has at least 5-fold fewer transcription regulatory sequences.
3. The synthetic nucleic acid molecule of claim 1 wherein the codon composition of the synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 35% of the codons.
4. The synthetic nucleic acid molecule of claim 1 wherein the codon composition of the synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 45% of the codons.
5. The synthetic nucleic acid molecule of claim 1 wherein the codon composition of the synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 55% of the codons.

6. The synthetic nucleic acid molecule of claim 1 wherein the majority of codons which differ are ones that are preferred codons of a desired host cell.
7. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule encodes a reporter molecule.
8. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule encodes a selectable marker protein.
9. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule encodes a luciferase.
10. The synthetic nucleic acid molecule of claim 9 wherein the wild type nucleic acid sequence encodes a *Renilla* luciferase.
11. The synthetic nucleic acid molecule of claim 9 wherein the wild type nucleic acid sequence encodes a beetle luciferase.
12. The synthetic nucleic acid molecule of claim 11 wherein the synthetic nucleic acid molecule encodes the amino acid valine at position 224.
13. The synthetic nucleic acid molecule of claim 11 wherein the synthetic nucleic acid molecule encodes the amino acid histidine at position 224, histidine at position 247, isoleucine at position 346, glutamine at position 348, or any combination thereof.
14. The synthetic nucleic acid molecule of claim 1 wherein the majority of codons which differ in the synthetic nucleic acid molecule are those which are employed more frequently in mammals.

15. The synthetic nucleic acid molecule of claim 1 wherein the majority of codons which differ in the synthetic nucleic acid molecule are those which are preferred codons in humans.
16. The synthetic nucleic acid molecule of claim 1 wherein the majority of codons which differ in the synthetic nucleic acid molecule are those which are preferred codons in plants.
17. The synthetic nucleic acid molecule of claim 9 wherein the synthetic nucleic acid molecule comprises SEQ ID NO:21 (Rlucver2) or SEQ ID NO:22 (Rluc-final).
18. The synthetic nucleic acid molecule of claim 9 wherein the synthetic nucleic acid molecule comprises SEQ ID NO:7 (GRver5), SEQ ID NO:8 (GRver6), SEQ ID NO:9 (GRver5.1), or SEQ ID NO:297 (GRver5.1).
19. The synthetic nucleic acid molecule of claim 9 wherein the synthetic nucleic acid molecule comprises SEQ ID NO:14 (RDver5), SEQ ID NO:15 (RDver7), SEQ ID NO:16 (RDver5.1), SEQ ID NO:299 (RDver5.1), SEQ ID NO:17 (RDver5.2), SEQ ID NO:18 (RD156-1H9) or SEQ ID NO:301 (RD156-1H9).
20. The synthetic nucleic acid molecule of claim 15 wherein the majority of codons which differ are the human codons CGC, CTG, TCT, AGC, ACC, CCA, CCT, GCC, GGC, GTG, ATC, ATT, AAG, AAC, CAG, CAC, GAG, GAC, TAC, TGC and TTC.
21. The synthetic nucleic acid molecule of claim 15 wherein the majority of codons which differ are the human codons CGC, CTG, TCT, ACC, CCA,

GCC, GGC, GTC, and ATC or codons CGT, TTG, AGC, ACT, CCT, GCT, GGT, GTG and ATT.

22. The synthetic nucleic acid molecule of claim 16 wherein the majority of codons which differ are the plant codons CGC, CTT, TCT, TCC, ACC, CCA, CCT, GCT, GGA, GTG, ATC, ATT, AAG, AAC, CAA, CAC, GAG, GAC, TAC, TGC and TTC.
23. The synthetic nucleic acid molecule of claim 16 wherein the majority of codons which differ are the plant codons CGC, CTT, TCT, ACC, CCA, GTC, GGA, GTC, and ATC or codons CGT, TGG, AGC, ACT, CCT, GCC, GGT, GTG and ATT.
24. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule is expressed in a mammalian host cell at a level which is greater than that of the wild type nucleic acid sequence.
25. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of CTG or TTG leucine-encoding codons.
26. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of GTG or GTC valine-encoding codons.
27. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of GGC or GGT glycine-encoding codons.

28. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule an increased number of ATC or ATT isoleucine-encoding codons.
29. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of CCA or CCT proline-encoding codons.
30. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of CGC or CGT arginine-encoding codons.
31. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of AGC or TCT serine-encoding codons.
32. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of ACC or ACT threonine-encoding codons.
33. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule has an increased number of GCC or GCT alanine-encoding codons.
34. The synthetic nucleic acid molecule of claim 1 wherein the codons in the synthetic nucleic acid molecule which differ encode the same amino acids as the corresponding codons in the wild type nucleic acid sequence.
35. A plasmid comprising the synthetic nucleic acid molecule of claim 1.

36. An expression vector comprising the synthetic nucleic acid molecule of claim 1 linked to a promoter functional in a cell.
37. The expression vector of claim 36 wherein the synthetic nucleic acid molecule is operatively linked to a Kozak consensus sequence.
38. The expression vector of claim 36 wherein the promoter is functional in a mammalian cell.
39. The expression vector of claim 36 wherein the promoter is functional in a human cell.
40. The expression vector of claim 36 wherein the promoter is functional in a plant cell.
41. The expression vector of claim 36 wherein the expression vector further comprises a multiple cloning site.
42. The expression vector of claim 41 wherein the expression vector comprises a multiple cloning site positioned between the promoter and the synthetic nucleic acid molecule.
43. The expression vector of claim 41 wherein the expression vector comprises a multiple cloning site positioned downstream from the synthetic nucleic acid molecule.
44. A host cell comprising the expression vector of claim 36.
45. A reporter gene expression kit comprising, in suitable container means, the expression vector of claim 36.

46. An isolated polypeptide encoded by SEQ ID NO:9 (GRver5.1) or SEQ ID NO:18 (RD156-1H9).
47. A polynucleotide which hybridizes under stringent hybridization conditions to SEQ ID NO:22 (Rluc-final), SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297 (GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof.
48. A method to prepare a synthetic nucleic acid molecule comprising an open reading frame, comprising:
- a) altering a plurality of transcription regulatory sequences in a parent nucleic acid sequence which encodes a polypeptide having at least 100 amino acids to yield a synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to the parent nucleic acid sequence, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites, enhancer sequences and promoter sequences; and
 - b) altering greater than 25% of the codons in the synthetic nucleic acid sequence which has a decreased number of transcription regulatory sequences to yield a further synthetic nucleic acid molecule, wherein the codons which are altered do not result in an increased number of transcription regulatory sequences, wherein the further synthetic nucleic acid molecule encodes a polypeptide with at least 85% amino acid sequence identity to the polypeptide encoded by the parent nucleic acid sequence.
49. A method to prepare a synthetic nucleic acid molecule comprising an open reading frame, comprising:

a) altering greater than 25% of the codons in a parent nucleic acid sequence which encodes a polypeptide having at least 100 amino acids to yield a codon-altered synthetic nucleic acid molecule, and

b) altering a plurality of transcription regulatory sequences in the codon-altered synthetic nucleic acid molecule to yield a further synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to a synthetic nucleic acid molecule with a random selection of codons at the codons which differ, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites, enhancer sequences and promoter sequences, and wherein the further synthetic nucleic acid molecule encodes a polypeptide with at least 85% amino acid sequence identity to the polypeptide encoded by the parent nucleic acid sequence.

50. The method of claim 48 or 49 wherein the parent nucleic acid sequence encodes a reporter molecule.
51. The method of claim 48 or 49 wherein the parent nucleic acid sequence encodes a luciferase.
52. The method of claim 48 or 49 wherein the synthetic nucleic acid molecule hybridizes under medium stringency hybridization conditions to the parent nucleic acid sequence.
53. The method of claim 48 or 49 wherein the codons which are altered encode the same amino acid as the corresponding codons in the parent nucleic acid sequence.
54. A synthetic nucleic acid molecule which is the further synthetic nucleic acid molecule prepared by the method of claim 48 or 49.

55. A method for preparing at least two synthetic nucleic acid molecules which are codon distinct versions of a parent nucleic acid sequence which encodes a polypeptide, comprising:
- a) altering a parent nucleic acid sequence to yield a synthetic nucleic acid molecule having an increased number of a first plurality of codons that are employed more frequently in a selected host cell relative to the number of those codons in the parent nucleic acid sequence; and
 - b) altering the parent nucleic acid sequence to yield a further synthetic nucleic acid molecule having an increased number of a second plurality of codons that are employed more frequently in the host cell relative to the number of those codons in the parent nucleic acid sequence, wherein the first plurality of codons is different than the second plurality of codons, and wherein the synthetic and the further synthetic nucleic acid molecules encode the same polypeptide.
56. The method of claim 55 further comprising altering a plurality of transcription regulatory sequences in the synthetic nucleic acid molecule, the further synthetic nucleic acid molecule, or both, to yield at least one yet further synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to the synthetic nucleic acid molecule, the further synthetic nucleic acid molecule, or both.
57. The method of claim 55 further comprising altering at least one codon in the first synthetic sequence to yield a first modified synthetic sequence which encodes a polypeptide with at least one amino acid substitution relative to the polypeptide encoded by the first synthetic nucleic acid sequence.

58. The method of claim 56 further comprising altering at least one codon in the second synthetic sequence to yield a second modified synthetic sequence which encodes a polypeptide with at least one amino acid substitution relative to the polypeptide encoded by the first synthetic nucleic acid sequence.
59. The method of claim 55 wherein the synthetic sequences encode a luciferase.
60. The synthetic nucleic acid molecule of claim 1 wherein the synthetic nucleic acid molecule is expressed at a level which is at least 110% of that of the wild type nucleic acid sequence in a cell or cell extract under identical conditions.
61. The synthetic nucleic acid molecule of claim 1 wherein the polypeptide encoded by the synthetic nucleic acid molecule has at least 90% contiguous sequence identity to the polypeptide encoded by the wild type nucleic acid sequence.
62. The synthetic nucleic acid molecule of claim 1 wherein the polypeptide encoded by the synthetic nucleic acid molecule is identical in amino acid sequence to the polypeptide encoded by the wild type nucleic acid sequence.
63. A vector comprising a synthetic nucleic acid molecule having at least 3-fold fewer transcriptional regulatory sequences relative to a vector comprising a parent nucleic acid sequence, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites and promoter sequences.
64. The vector of claim 63 wherein the synthetic nucleic acid molecule does not encode a polypeptide.

65. The method of claim 48 or 49 further comprising altering the further synthetic nucleic acid molecule to encode a polypeptide having at least one amino acid substitution relative to the polypeptide encoded by the parent nucleic acid sequence.
66. The method of claim 48 or 49 wherein the altering of transcription regulatory sequences does not introduce amino acid substitutions to the polypeptide encoded by the synthetic nucleic acid molecule.

004280" 90254950